

本周工作：

1. 本周继续边界项目：

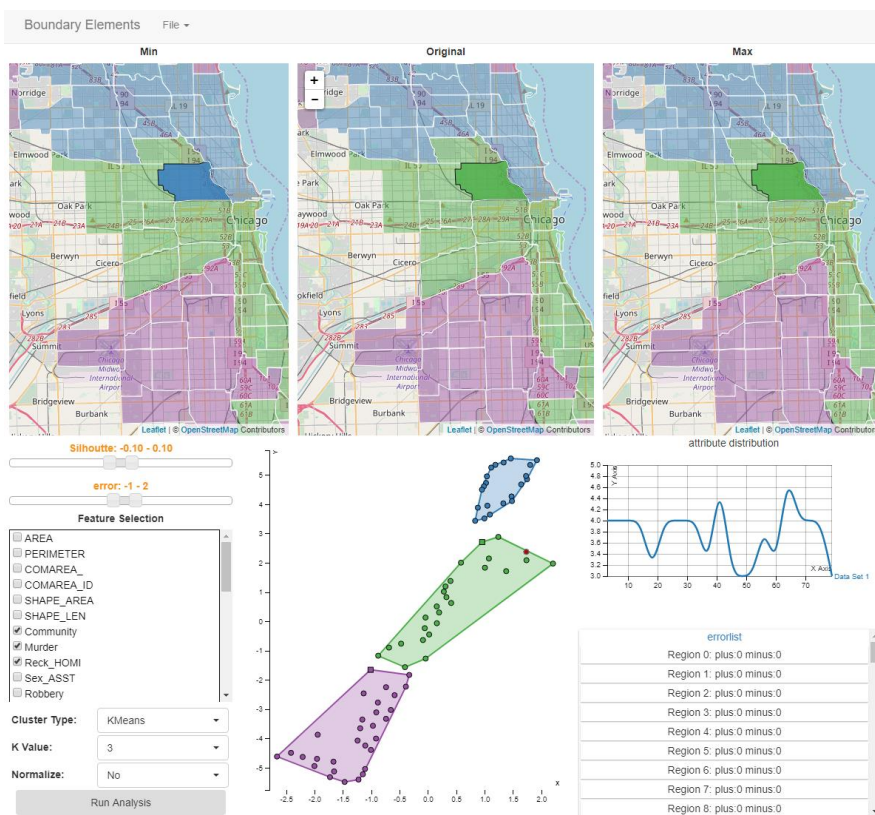
完成了代码到目前为止讨论的版本，并实现 tsne 和 kmeans 算法的参数交互。

本周和 ross 及 yafeng 讨论的任务有：

当地图聚类时，某一块地区的某一个参数可能会有一定的误差，当这个误差范围变化时，哪些地区对聚类结果的影响较大，他们分别在哪里。

实现这一目标的难点有：值变化后的 kmeans 需要稳定的聚类结果，聚类后的结果需要对比。交互的联动。

本周实现了这三部分的代码工作。其中 kmeans 的稳定用到的是机器学习的一个库，mlkmeans，每次迭代聚类中心，起始的聚类中心相同。本项目核心代码基本完成。下周周一和 ross 约好继续讨论下一步。



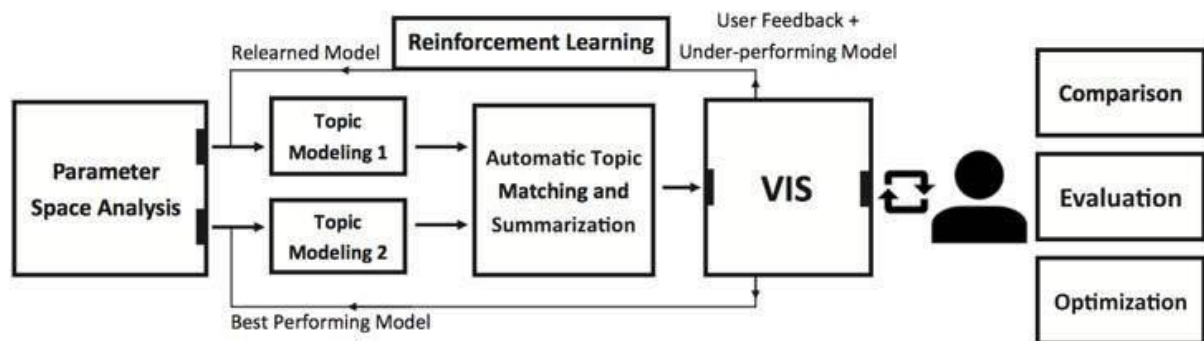
2. 与明洁讨论集群

集群的工作还是主要集中在数据的流程方面，我们讨论了关于用可视化控制集群分配任务，控制集群处理的中间结果并优化算法的方法，但是明杰说难度很大，集群不好实现。

3. 论文：

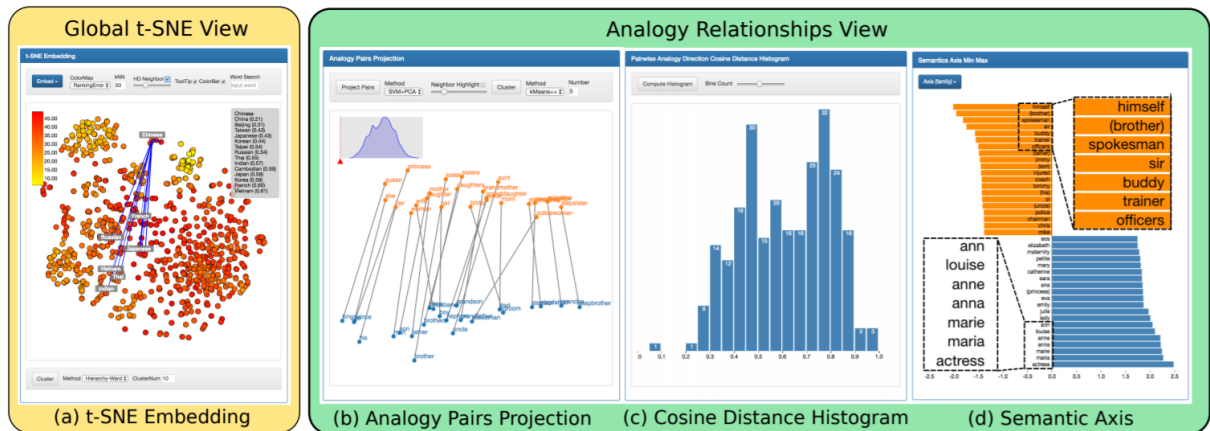
Progressive Learning of Topic Modeling Parameters: A Visual Analytics Framework

虽然主题模型的抽取有利于用户了解数据内容，完成数据分类等任务，但是还是太抽象，本文增强了主题模型，对于同一个语料库，本文利用两个模型抽取算法，然后相互匹配得到相匹配的主题，相当于通过多角度的观测，得到主题的更加全面的语义。同时提供对比，评估优化的方法，给用户分析交互。



Visual Exploration of Semantic Relationships in Neural Word Embeddings

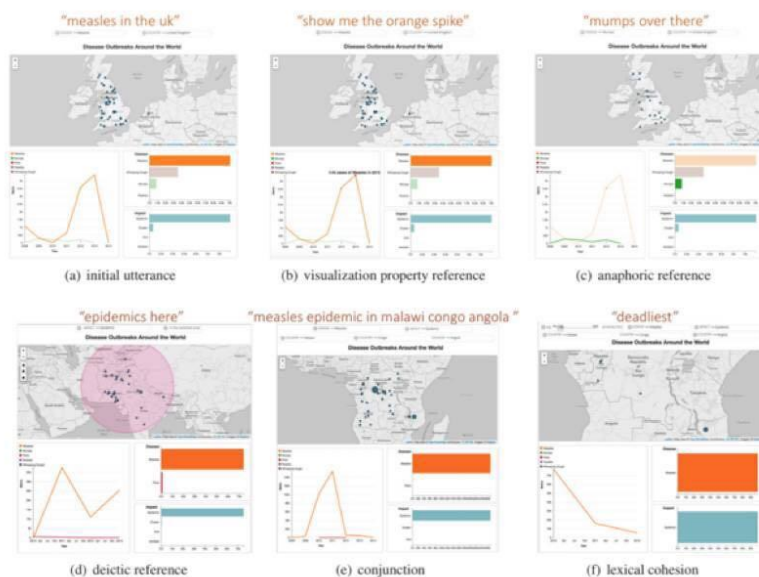
利用词嵌入模型进行的向量分析方法流程已经成为自然语言处理的关键流程。然而，这些空间的结构和性质难以了解。Tsne 和 PCA 的投影方法可以用来在 2 维向量中分析对象之间的关系，评估总体结构和线性关系。本文在二维投影中，利用 x 轴编码类比关系中两个概念的区别，Y 轴编码不同的次对之间的区别，然后利用投影方法计算各个对象的投影位置，从而提出了一种可用于语义和句法类比的新的嵌入技术.同时利用统计视图和 tsne 视图帮助理解整体属性和细节分析，语义信息等。



Applying Pragmatics Principles for Interaction with Visual Analytics.

本文的结论使我有启发，当用户想要分析任务时，其实它往往更加关心展示的数据，而不是如何操作交互，使得能够分析。交互只是用来分析的手段，当然是越简单有效越好。语义轨迹的查询也可以扯扯这方面，从简化查询输入谈谈。本位利用自然语言的可视化界面，并给予评估，方便，快捷，吸引用户

Enamul Hoque, Vidya Setlur, Melanie Tory, and Isaac Dykeman



边界	程序，讨论	12 月初
语义轨迹	规划交互	12 月底